```
Suggestion with Lucene index
-----------------------------
    Shankar:
        analyzer,
        tokenizer,
        ...


    >> vector based storage.
    >> bag of word models.

    document: has words.

    [document as vector of inverted index.]

    d1: t11, t12, ..., t1k1
    d2: t21, t22, ..., t2k2
    d3

    dn
    ---------------------------
    doc/ terms
        t1  t2  t3  ... tK
    d1: s1  s2
    d2:

    where K = union (ki)

    ----------------------------
    score => normalized tf-idf.
    ----------------------------
    interverted index: to avoid sparse 0 zeros.
    ----------------------------

>> Tokenizer: return all the token terms for a document(strem of token).

d1: Shankar has large family with 3 teenage kids.

stream of ([shankar, has, large, family, with, 3, teenage, kids])

filters: stop words [a, an, the, has, numeric, ...]

filter takes input of (tokenStream)

Analyzer: StandardAnalyzer, WhiteSpace,....

    ngram analyzer: returns stream of ngrams from the document.

-------------------------------------------------------------------------
```